

Database Development Issues and Status

Rod Davidson, Sabre Systems, inc.

March 17, 2005

I've been contracted by NOAA Fisheries (NWFSC), Conservation Biology staff to develop a database structure, data migration tools, and preliminary application interfaces for Status-and-Trend monitoring data. My assignment was aimed at creating a shared data repository in a web-accessible Oracle environment. Over the past four months I have analyzed many required references, have met with a variety of staff members, and have designed and tested a prototype database and migration software towards satisfying this contract.

To expedite the development of this data solution, it has been necessary for me to minimize meeting involvement and extended discussions with many subject-matter specialists. I do, however, want to keep communication with the stakeholders, so offer the following discussion of the project status.

PROTOCOL Definition Issues

- Ongoing transition/refinement of Protocol-Builder – The contract to build a common data structure was assigned while the definition of common protocol definitions was still underway. In fact, at the time of this writing, the definition of common protocol specifications is still evolving. While I have made considerable progress in building the data structure during this time, it is evident that modifications to the structure will likely be necessary once the protocol definitions become more firm. I'm concerned that time may not be available within this contract to accomplish the modifications once they are defined.
- Protocol-versus-Event data items – A number of data items identified in the modeling process as event characteristics are also included as part of each protocol's definition. They are becoming accepted as protocol characteristics in order to satisfy an objective of the Protocol Builder form generation. However, in the implementation of the data model I am not treating them as protocol characteristics because they are the result of a monitoring event. Rather, I've structured them for the following definitions:
 - A monitoring site is located along a reach (the WHERE).
 - A monitoring event is conducted by a team at a monitoring site (WHO and WHEN). Each monitoring event employs one or more protocols to obtain measurements.
 - The protocol instructs the team which measurements or observations are to be taken and by what procedures. (WHAT and HOW)
- Attribute Codes – The construction of attribute code sets in the Protocol Dictionary appears to be incomplete because, of the 1245 attributes, there are 984 attributes assigned to code set '0'. So, there are 260 attributes that are set to a code set. (The code set '0' appears to consist of "B=Boat, BP=Backpack, BT=Bank/Tow, and S-Seine" so is not likely to apply to 984 attributes.) When the code set definitions are completed, we will update the lookup tables in this package.

SITE and EVENT Architecture Solutions

I want to explain some of the fundamental organization of the data architecture since there has been much discussion about Monitoring Events and Monitoring Sites. For the organization of these objects, I tried to implement some simple rules and definitions of these data objects.

- A Monitoring Site: is the place where an event occurs. For our purposes, sites can have a variety of geometries; including Reaches, Transects, X-Sites, Reach Segments, etc. Each of these geometries may have an event occur at the site where it is located. The geometry of the site usually determines what measurements that the event obtains.
- A Monitoring Event: is the activity of a team of personnel (sponsored by an organization) making a visit to a site on a specific date (and, perhaps, time) to obtain some measurements. Again, the measurements are often determined by what site "geometry" they are visiting.
- The measurements taken at a site, by a team, on a date, are dictated the site geometry and are meaningless without their context.

I hope that these three rules make the model easier to comprehend.

DATA MIGRATION Issues

- Large-Woody Debris Protocol – Although the Monitoring Strategy specifies (and the protocol definition implements) particular categories of large woody debris, it appears that the Wenatchee field data was counted using different criteria. For example:

Definition in Protocol Dictionary

Criteria used in Wenatchee Field Data

Len_1_5m	= Count of debris measuring between 1.5 - 5m	10_Short	= Estimated number of large wood pieces that are 10-15 cm diameter and 1-3 m long.
Len_5_15m	= Count of debris measuring between 5 - 15m	10_Med	= Estimated number of large wood pieces that are 10-15 cm diameter and 3-6 m long.
Len>15m	= Count of debris measuring Greater than 15m	10_Long	= Estimated number of large wood pieces that are 10-15 cm diameter and >6 m long.
None		15_Short	= Estimated number of large wood pieces that are 15-30 cm diameter and 1-3 m long.
None		15_Med	= Estimated number of large wood pieces that are 15-30 cm diameter and 3-6 m long.
None		15_Long	= Estimated number of large wood pieces that are 15-30 cm diameter and >6 m long.
None		30_Short	= Estimated number of large wood pieces that are >30 cm diameter and 1-3 m long.
None		30_Med	= Estimated number of large wood pieces that are >30 cm diameter and 3-6 m long.
None		30_Long	= Estimated number of large wood pieces that are >30 cm diameter and >6 m long.

- The following measures are found in the Wenatchee Field Data. Their corresponding definitions in the Protocol Dictionary are not obvious:

Channel Condition Protocols

Subject Area	Protocol Name	Attribute Name	Attribute Description
Channel Condition	Phab-ChanCS-BankMeasure	FlagBank	notes taken in the field that qualify conditions that don't strictly meet categories/measurement conditions appropriate for other Attributes within this Protocol. A value here tells user which of the comments to look at in the comment attribute for more information.
Channel Condition	Phab-ChanCS-BankMeasure	BankfullDepth	Average depth of channel at Bankfull
Channel Condition	Phab-ChanCS-BankMeasure	FloodproneDepth	Two times the bankfull depth.
Channel Condition	Phab-ChanCS-BankMeasure	FloodproneWidth	Width of the channel at the floodprone width.

Mike Ward's Comments

I suggest adding this attribute name and I suggest possibly converting all the "FLAG" into more specific attribute names. But I'm not sure; see notes below for Flag within ProtocolName=HQEmbeddedness, this is another interesting database challenge.

I suggest adding this attribute name because I don't see it elsewhere in channel condition. This is by definition supposed to be 2XBankfullDepth, but it is pencilled in in the field, so maybe we should program a check on this value. I suggest adding this attribute name because I don't see it elsewhere in channel condition.

I suggest adding this attribute name because I don't see it elsewhere in channel condition.

In addition to the above, the following attributes are also found in Channel Condition protocols for the Wenatchee Field Data but cannot be found in the Protocol Dictionary:
ENDPOINTS, RUN_M, EYE_HGT, LVL_HGT, RISE_CM, RADIANS, NORTHING, EASTING

Mike's Comments: **I do not believe that these attributes were included in the original data dictionary, though they are part of the Rosgen protocols. These may be kinda like SUBJECT AREA = Channel Condition, PROTOCOL = Phab-SlopeBearing, but I'm not sure where those SUBJECT AREA/PROTOCOL attributes are intended to be used. I'm guessing we need to define new attributes to capture this measurement.

Habitat Quality Protocols

notes taken in the field that qualify conditions that don't strictly meet categories available from SizeClassCode. A value here tells user which of the comments to look at in the comment attribute for more information.

Habitat Quality HQ-Embeddedness FlagSubstrate

Habitat Quality Phab-ChanCS-FishCover BrushWoodyDebris

Habitat Quality Phab-ChanCS-FishCover Flag

Percentage of 10 m length of bank (at bankfull stage) at each transect that appears unstable due to breakdown, slumping, cracking, or bare/steep surfaces.

Habitat Quality Phab-ChanCS-FishCover UnstableBanks

I strongly recommend renaming this to "Brush" instead of "BrushWoodyDebris"

I don't know what to do here. For each attribute in this protocol, WDOE's field form has a cell for "Flag" but gives them all different names in the Access database (E.g. Algae_F xwalks into FlagAlgae [just like FlagSubstrate]). This seems very cumbersome from a data mgt perspective. It will depend how you decide to deal with Flags as to how we should go.

Need to add this attribute

- Standard Data Sources – Field data received to date has been in multiple formats which can be accommodated, but require considerable investment in software development. For example, the Habitat Quality data is in an MS-Access database while the steelhead redds data is in an Excel spreadsheet. As such data is received considerable expense can be avoided if future submissions of the same topical data are provided in these original formats. The migration software will be more easily reusable.
- Definition of Stream LLID (Latitude Longitude Identifier) – The standard definition for LLID specifies that it be a 13-digit number with 7-digits for longitude and 6-digits for latitude. However in the Wenatchee Habitat data set the numbers don't appear to follow that definition. I will need assistance in devising a method for interpreting the LLIDs for the Wenatchee habitat data?

DESIGN ENHANCEMENTS

- Fully defined PLSS locations – In the Public Land Survey System the Townships and Ranges (which define section groups) need to be identified with a Principal Meridian for proper interpretation. Although Washington and Oregon surveys are based on the Willamette Meridian, Idaho surveys are based on the Boise Meridian. (And, Alaska has multiple meridians for PLSS.) The mixing of these states into one data repository will require the meridian designation along with Township and Range identifiers. I have included Principal Meridian in the database design.
- Units-of-Measure – The units that a measurement is in when recorded are identified in most protocols. In addition, I have designed into this model a unit-conversion function. This will enable the units-of-measure to be dynamically converted to other relevant units for specialized reporting purposes. If a set of measurements is taken in non-standard units, this also enables dynamic conversion, at the time of output, to a consistent unit-of-measure.